

Maximal entropy distribution functions from generalized Rényi entropy

György Steinbrecher

Association EURATOM-MEdC, University of Craiova,
A. I. Cuza 13, 200585 Craiova, Romania.
Email: gyorgy.steinbrecher@gmail.com

Giorgio Sonnino

Department of Theoretical Physics and Mathematics
Université Libre de Bruxelles (ULB)
Bvd de Triomphe, Campus Plaine CP 231, 1050 Brussels, Belgium.
&

Royal Military School (RMS)

Av. de la Renaissance 30, 1000 Brussels, Belgium.
E-mail: gsonnino@ulb.ac.be

Nicolae Pometescu

Association EURATOM-MEdC, University of Craiova,
A. I. Cuza 13, 200585 Craiova, Romania.
Email npomet@yahoo.com

November 22, 2016

Abstract

New class of reference distribution functions for numerical approximation of the solution of the Fokker-Planck equations associated to the charged particle dynamics in tokamak are studied. The reference distribution functions are obtained by maximization of the generalized Rényi entropy under scale-invariant restrictions. Explicit analytic form, with algebraic decay, that is a generalization of the previous distribution with exponential tails was derived.

1 Introduction

The two independently discovered generalizations of the classical Boltzmann-Shannon Entropy (BSE) [1], by Constantino Tsallis [2]-[5] respectively by Alfréd Rényi [6]-[9] found multiple applications [4], [5], [10]-[12]. It is a challenging

problem to explain the successes of two similar generalizations of the BSE. The mathematical "naturalness", of these generalizations was explained in the framework of Lebesgue L^p spaces defined over general measure space, in the sense that both Tsallis entropy (TE) and the additive Rényi entropy (RE) of order p contains the same L^p norm for $p > 1$ respectively the L^p pseudonorm for $0 < p < 1$ [13], [14], [15], [16]. A natural generalization of the Rényi entropy (GRE), preserving the its additivity and geometrical interpretation as a distance in the functions vector space, where the probability density function is, was given in [16], [17]. It was proven [17] that by imposing a suitable stabilizing conditions, the TE, RE and GRE are numerically stable for a large range of the parameters, they are more stable compared to BSE in the sense that for the stability of BSE more stabilizing conditions must be imposed [17]. The numerical stability of TE, RE and GRE are related to logarithmic convexity property [17]. It was proven in that both the TE, RE, GRE, contains a functional that has good category theoretic properties. The GRE has an interesting application in the study of the complex dynamical systems [25]. In a suitable limit the GRE became the RE, that whose limiting value is the BSE, so the GRE appears as a natural generalization of the RE, for characterization of the singularity or asymptotic behavior of the multivariate PDF [16], [17], or in the characterization of discrete probability distributions where the set of states is a Cartesian product (the probabilities has multiple indices) [19].

The GRE is a Liapunov functional for a large class of dynamical systems driven by stochastic perturbations [16]. Consequently it is meaningful the study of PDF that realize the maximizes the generalized entropies. In contrast to the case of BS, RE, TE, in the case of GRE, even in the simplest case when the total phase space is a Cartesian product of two smaller spaces, the stationarity condition is expressed by a more complicated functional equation [16], whose solutions has complex algebraic decays. The maximal generalized entropy (GMax-Ent) PDF's, whose study is the object of this article, are interesting because in a series of our previous works [20]-[23] we proved that the reference distribution function (RDF) for charged particle distribution in tokamak can be obtained by maximization of the BSE, subject to scale invariant, algebraically the simplest, restrictions. In order to obtain a better approximation of the PDF of the charged particles in tokamak, in this work we enlarge the family of RDF obtained previously by MaxEnt principle with scale invariant restrictions, by considering RDF's obtained from the maximization of the Generalized Rényi entropy.

Reference particle density distribution functions are useful in the numerical solutions of Fokker-Planck or gyrokinetic equations, that describe charged particle dynamics in tokamak.

Simplest MaxgEnt distribution functions were studied in [16]. In this article we explore systematically the class of GRE that appears when the phase space is a Cartesian product of $N = 3$ sub spaces. We use the general formalism of the Lebesgue integration theory that allows to have an unified formalism for both discrete and continuous distributions in finite as well as infinite dimensional spaces of stochastic processes. In the our formalism the Rényi divergence

appears as a Rényi entropy for a suitable chosen measure [19].

2 The framework, Shannon, Rényi and Tsallis entropies

For starting the discussion about various measures of the information, we need to specify some exact framework (see [17]). We consider a standard measure space (Ω, \mathcal{A}, m) where Ω is the phase space, \mathcal{A} is the σ -algebra of the observable events, m is the measure (that can be discrete, continuous, finite or σ -finite).

In the classical definitions with discrete, finite or denumerable probability space, the measure m is a the counting measure, invariant under permutation group. In many applications when Ω is a continuum, the measure m is invariant under physical or geometrical symmetries. In this article the probability measures

$$\mathcal{A} \ni A \rightarrow p(A) \in [0, 1]$$

defined on (Ω, \mathcal{A}) are continuous with respect to measure m , so by Radon-Nicodim theorem we have

$$p(A) = \int_A \rho(x) dm(x); \quad A \subset \Omega \quad (1)$$

where $\rho(x)$ is the probability density function. With the previous notations the Boltzmann-Gibbs-Shannon entropy has the following form

$$S_{cl}[\rho] = - \int_{\Omega} \rho(x) \log [\rho(x)] dm(x) \quad (2)$$

In the definitions of the A. Rényi [6] respectively by C. Tsallis [2], [3] entropies we encounter the same metric object in the $L^p(\Omega, dm)$ spaces [13], [14], [15]. For details see ref. [16]. Consequently, with the notations

$$\|\rho\|_p = \left[\int_{\Omega} [\rho(x)]^p dm(x) \right]^{\frac{1}{p}}; \quad p \geq 1 \quad (3)$$

$$N_p[\rho] = \int_{\Omega} [\rho(x)]^p dm(x); \quad 0 < p \leq 1 \quad (4)$$

the entropies of A. Rényi [6] $S_{R,q}$ respectively by C. Tsallis [2], [3] $S_{T,q}$ can be

expressed as follows

$$S_{R,q}[\rho] = \frac{q}{1-q} \log \|\rho\|_q; q > 1 \quad (5)$$

$$S_{R,q}[\rho] = \frac{1}{1-q} \log N_q[\rho]; 0 < q < 1 \quad (6)$$

$$S_{T,q}[\rho] = \frac{1}{1-q} \left[1 - \|\rho\|_q^q \right]; q > 1 \quad (7)$$

$$S_{T,q}[\rho] = \frac{1}{1-q} \{1 - N_q[\rho]\}; 0 < q < 1 \quad (8)$$

In this formalism the Rényi divergence can be expressed as Rényi entropy with suitable chosen measure $dm(\mathbf{x})$ [19].

Remark 1 *Observe from previous formalism in the case of discrete probability distribution the original definitions of the Rényi or Tsallis entropies results [6], [2].*

$$S_{R,q}[\rho] = \frac{1}{1-q} \log \sum_k p_k^q \quad (9)$$

$$S_{T,q}[\rho] = \frac{1}{1-q} \left[1 - \sum_k p_k^q \right] \quad (10)$$

From the previous definitions it is clear that the Rényi and Tsallis entropies are related to the geometric properties Lebesgue spaces $L^p(\Omega, dm)$, their norms or pseudo norms.

3 The generalized Rényi entropies (GRE).

3.1 Definitions and notations.

We follow the same approach that from ref.[16], [17]. We will define the Generalized Rényi entropies by using the results on Banach spaces with the anisotropic norm, exposed in ref.[18]. In the following we will restrict our discussions to the set of parameters that define the GRE, when a) The integrals that appears in the definition can be interpreted like distance in a suitable function space and b) The formula for entropy can be related to convexity or concavity properties of some functional, in the subspace of non negative density functions. Consequently we will define only two class of distance functionals and entropies, in analogy to the functionals $S_{p_y, p_z}^{(1)}[\rho]$ and $S_{q_y, q_z}^{(2)}[\rho]$ defined in ref.[16].

Consider that the measure space (Ω, \mathcal{A}, m) has the following product structure. The phase space Ω is split in 3 subspaces

$$\Omega = \Omega_1 \times \Omega_2 \times \Omega_3 \quad (11)$$

That means that the argument \mathbf{x} of probability density function can be represented as $\mathbf{x} = \{x_1, x_2, x_3\}$, so

$$\rho(\mathbf{x}) = \rho(x_1, x_2, x_3) \quad (12)$$

with $x_k \in \Omega_k$. We mention also that in general the component spaces Ω_k has the structure of \mathbf{R}^n or more general infinite dimensional measure space. Each of the spaces Ω_k has their σ -algebra \mathcal{A}_k . The σ -algebra \mathcal{A} , that contains subsets of $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_N$ is defined as a tensor product: it is the largest σ -algebra on Ω such that all of the projections $\Omega \xrightarrow{p_k} \Omega_k$ are measurable.

The measure m is also Factorizable:

$$dm(\mathbf{x}) = dm(x_1, x_2, x_3) = \prod_{j=1}^3 dm_j(x_j) \quad (13)$$

where the measures m_k are defined on the σ -algebras \mathcal{A}_k .

In other words, the measure space (Ω, \mathcal{A}, m) is the tensor product

$$(\Omega, \mathcal{A}, m) = \bigotimes_{j=1}^3 (\Omega_j, \mathcal{A}_j, m_j) \quad (14)$$

The elementary probability $dP(\mathbf{x})$ is given by

$$dP(\mathbf{x}) = \rho(x_1, x_2, x_3) dm(\mathbf{x}) \quad (15)$$

where $dm(\mathbf{x})$ is given by Eq.(13).

Consider a vector $\mathbf{p} = \{p_1, p_2, p_3\}$ of real numbers with $p_k \geq 1$. According to Ref.[18], in close analogy to Ref.[16] (where the particular case $N = 2$ was studied) and [17], we define recursively the norm (depending on the measure m) $\|\rho\|_{\mathbf{p}, m}$ as follows

$$\rho_2(x_1, x_2) := \left[\int_{\Omega_N} [\rho(x_1, x_2, x_3)]^{p_3} dm_3(x_3) \right]^{1/p_3} \quad (16)$$

$$\rho_1(x_1) := \left[\int_{\Omega_2} [\rho_2(x_1, x_2)]^{p_2} dm_2(x_2) \right]^{1/p_2} \quad (17)$$

$$\|\rho\|_{\mathbf{p}, m} := \left[\int_{\Omega_1} [\rho_1(x_1)]^{p_1} dm_1(x_1) \right]^{1/p_1} \quad (18)$$

In analogy with Eqs.(3, 5) and Ref.[16] we define the GRE, with respect to the measure m

$$S_{\mathbf{p}}^{(1)}[\rho, m] = \frac{p_1}{1 - p_3} \log \|\rho\|_{\mathbf{p}, m}; p_i > 1 \quad (19)$$

Observe that the anisotropic norm function $\rho \rightarrow \|\rho\|_{\mathbf{p},m}$ is convex and satisfies the axioms of norm. The corresponding normed vector space is complete, i. e. it is a Banach space. See ref.[18]. There is another range of parameters that generalize the Rényi entropy corresponding to Eqs.(4, 6). Consider a vector $\mathbf{q} = \{q_1, q_2, q_3\}$ of real numbers with $0 < q_k \leq 1$. In analogy to Eqs.(16-18) we define recursively [17]

$$\rho'_2(x_1, x_2) := \int_{\Omega_N} [\rho(x_1, x_2, x_3)]^{q_3} dm_3(x_3). \quad (20)$$

$$\rho'_1(\mathbf{x}_1) := \int_{\Omega_2} [\rho'_2(x_1, x_2)]^{q_2} dm_2(x_2) \quad (21)$$

$$N[\rho]_{\mathbf{q},m} := \int_{\Omega_1} [\rho'_1(x_1)]^{q_1} dm_1(x_1) \quad (22)$$

Observe that the mapping $\rho \rightarrow N[\rho]_{\mathbf{p},m}$ defines a pseudonorm on the space of probability density functions. The map $\rho \rightarrow N[\rho]_{\mathbf{p},m}$ defines a concave function, in the subset of physically admissible PDF's, when $\rho(x_1, x_2, x_3) \geq 0$. The GRE will be defined in analogy to Eqs.(4, 6) and to the case $N = 2$ from ref. [16], [17]

$$S_{\mathbf{q}}^{(2)}[\rho, m] = \frac{1}{1 - q_3} \log N[\rho]_{\mathbf{q},m}; 0 < q_i < 1 \quad (23)$$

For simplification of the notations we will use the extrapolated for of the Eq.(23) also to the range of parameters $\{q_1, q_2, q_3\}$ that allows to relate $S_{\mathbf{q}}^{(2)}[\rho, m]$ to $S_{\mathbf{p}}^{(1)}[\rho, m]$. We obtain

$$S_{\mathbf{q}}^{(2)}[\rho, m] = S_{\mathbf{p}}^{(1)}[\rho, m] \quad (24)$$

$$N[\rho]_{\mathbf{q},m} = \left[\|\rho\|_{\mathbf{p},m} \right]^{p_1} \quad (25)$$

when p_i and q_i are related as follows

$$q_3 = p_3 \quad (26)$$

$$q_2 = \frac{p_2}{p_3} \quad (27)$$

$$q_1 = \frac{p_1}{p_2} \quad (28)$$

Remark 2 *The algebraic equations associated to maximal entropy problem are very complicated in the general case, nevertheless from the convexity or concavity properties we have some informations. According to Eqs.(24-28), we are in the domain when $\rho \rightarrow \|\rho\|_{\mathbf{p},m}$ is a convex functional when*

$$p_k = \prod_{j=k}^3 q_j \geq 1; 1 \leq k \leq 3 \quad (29)$$

In this case the problem of maximal entropy with linear restriction is equivalent to minimization of a positive convex function and has unique solution. In the domain $0 < q_k < 1$, where the map $\rho \rightarrow N[\rho]_{\mathbf{q},m}$ is a concave function, the maxent problem is equivalent with the maximization of a concave function with linear restriction. If the solution exists it is unique.

3.2 Particular cases.

In the following we will not omit the measure, when no confusion arise: $\|\rho\|_{\mathbf{p},m} := \|\rho\|_{\mathbf{p}}$; $N[\rho]_{\mathbf{q},m} := N[\rho]_{\mathbf{q}}$; $S_{\mathbf{p}}^{(a)}[\rho, m] := S_{\mathbf{p}}^{(a)}[\rho]$.

In the particular case when $p_1 = p_2 = p_3 > 1$, or $q_1 = q_2 = 1$, $0 < q_3 < 1$ the GRE is equal to the classical Rényi entropy from Eqs.(5, 6):

$$S_{\mathbf{p}}^{(1)}[\rho] = S_{R,p_3}[\rho] = \frac{p_3}{1-p_3} \log \|\rho\|_{p_3} \quad (30)$$

$$\|\rho\|_{p_3} = \left[\int_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x})^{p_3} \right]^{1/p_3} \quad (31)$$

respectively

$$S_{\mathbf{q}}^{(2)}[\rho] = S_{R,q_3}[\rho] = \frac{1}{1-q_3} \log N[\rho]_{q_3} \quad (32)$$

$$N[\rho]_{q_3} = \int_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x})^{q_3} \quad (33)$$

We used the notation from Eqs.(3, 4, 17, 23). In particular when $p_N \searrow 1$ in Eqs. (30, 31), respectively when $q_1 = q_2 = q_{N-1} = 1$; $q_N \nearrow 1$ in Eqs.(32, 33), we obtain the classical Boltzmann-Shannon entropy

$$\lim_{p_1=p_2=p_3 \searrow 1} S_{\mathbf{p}}^{(1)}[\rho] = \lim_{q_1=q_2=1; q_3 \nearrow 1} S_{\mathbf{q}}^{(2)}[\rho] = - \int_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x}) \log \rho(\mathbf{x}) \quad (34)$$

Remark that the path to the limiting classical case is essential

$$\lim_{q_3 \nearrow 1} \lim_{q_1=q_2 \nearrow 1} S_{\mathbf{q}}^{(2)}[\rho] = - \int_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x}) \log \rho(\mathbf{x}) \quad (35)$$

while

$$\lim_{q_1=q_2 \nearrow 1} \lim_{q_3 \nearrow 1} S_{\mathbf{q}}^{(2)}[\rho] = \infty$$

3.3 Geometric properties of GRE

Despite for the physical application the previous definitions of the Rényi entropy and GRE are more advantageous, it is important to remark that in the our

geometric approach the basic objects are the norms defined in Eqs.(16-18) or pseudo norms defined in Eqs.(20-22).

In the case when $p_k \geq 1$, $\mathbf{p} = \{p_1, \dots, p_N\}$, (see [18]) the norm $\|\cdot\|_{\mathbf{p}}$ has the usual properties: for $a \in \mathbf{R}$ we have $\|a\rho\|_{\mathbf{p}} = |a| \|\rho\|_{\mathbf{p}}$, respectively

$$\|\rho_1 + \rho_2\|_{\mathbf{p}} \leq \|\rho_1\|_{\mathbf{p}} + \|\rho_2\|_{\mathbf{p}}. \quad (36)$$

In particular it follows the convexity of the mapping $\rho \rightarrow \|\rho\|_{\mathbf{p}}$: for $0 \leq \alpha \leq 1$ we have

$$\|\alpha\rho_1 + (1 - \alpha)\rho_2\|_{\mathbf{p}} \leq \alpha\|\rho_1\|_{\mathbf{p}} + (1 - \alpha)\|\rho_2\|_{\mathbf{p}} \quad (37)$$

In the case $0 < q_k \leq 1$, $\mathbf{q} = \{q_1, \dots, q_N\}$, the properties of the pseudo norms $N[\rho]_{\mathbf{q},m}$ defined in Eqs.(20-22) also allows geometrical interpretations. We have

$$N[\rho_1 + \rho_2]_{\mathbf{q}} \leq N[\rho_1]_{\mathbf{q}} + N[\rho_2]_{\mathbf{q}} \quad (38)$$

e. This can be seen by using the definition and the simple inequality $|x+y|^q \leq |x|^q + |y|^q$, with $0 < q \leq 1$. Instead of convexity we have the following concavity inequality **”in the first octant” only** : when $\rho_{1,2} \geq 0$

$$N[\alpha\rho_1 + (1 - \alpha)\rho_2]_{\mathbf{q}} \geq \alpha N[\rho_1]_{\mathbf{q}} + (1 - \alpha)N[\rho_2]_{\mathbf{q}} \quad (39)$$

that can be proven easily by using the concavity of the function $f(x) := x^q$ with $0 < q \leq 1$.

By defining, the distance function between distribution functions ρ_1 and ρ_2 in the infinite dimensional space of PDF's by $d(\rho_1, \rho_2) := \|\rho_1 - \rho_2\|_{\mathbf{p}}$ for $p_k \geq 1$ respectively $d(\rho_1, \rho_2) := N[\rho_1 - \rho_2]_{\mathbf{q}}$ for $0 < q_k \leq 1$, we have the triangle inequality

$$d(\rho_1, \rho_3) \leq d(\rho_1, \rho_2) + d(\rho_2, \rho_3) \quad (40)$$

that allows geometrical interpretation of GRE in term of distance in the functional space of admissible PDF's.

4 Maximal Generalized Entropy distributions

The main objective of the our work is to obtain a shortest derivation of the RDF for the charged particle distribution, studied in the previous works [20]-[23]. We mention that by using the classical MaxEnt principle in [23] we obtained new derivation of the RDF studied in [20]-[22].

We consider, in the framework of the notations from Eqs.(13, 14) the following problem: In the convex set \mathcal{K} of PDF, defined by the linear constraints and non-negativity condition

$$\int_{\Omega} dm(\mathbf{x})\rho(\mathbf{x})f_k(\mathbf{x}) = c_k; \quad 0 \leq k \leq M \quad (41)$$

$$\rho(\mathbf{x}) \geq 0 \quad (42)$$

find the PDF with maximal entropy a) $S_{\mathbf{p}}^{(1)}[\rho]$ with $p_k > 1$, or b) $S_{\mathbf{q}}^{(2)}[\rho]$, with $0 < q_k < 1$. In the constraints we included also the normalization: $f_0(\mathbf{x}) := 1$ and $c_0 = 1$. In the case a) according to Eq.(19) the MaxEnt problem is equivalent convex optimization problem: to find the point on \mathcal{K} that the closest, to the origin, **in the sense of the distance defined by the norm $\|\rho\|_{\mathbf{p},m}$** . Due to the convexity of the norm Eq.(37) and convexity of the set \mathcal{K} , the solution of the maxent problem is unique. The existence is related to converge problems. If it exists then the solution is unique., irrespective how complicated are the algebraic equations for Lagrange multipliers. In the case when the measure is finite, namely when $\int_{\Omega} dm(\mathbf{x}) < \infty$ and the functions $f_k(\mathbf{x})$ are all bounded, then it is possible to find a set of parameters c_k such that the MaxEnt problem has a solution

In the case b) from Eq.(23) results that the MaxEnt problem is to find the point $\rho \in \mathcal{K}$ with maximal value of $N[\rho]_{\mathbf{q}}$. We observe that the distance function $d(\rho, \mathbf{0}) := N[\rho]_{\mathbf{q}}$ unlike to the familiar Euclidian distance is concave for $\rho \geq 0$, not convex. From the convexity of \mathcal{K} and the concavity inequality Eq. (39) results that also in case b) the solution of the maxent problem exists.

4.1 Explicit MaxEnt distributions for $N = 3$

In the previous works [20]-[22] we obtained realistic PDF starting from classical or generalized MaxEnt principle with scale invariant restrictions. Apparently, by adding sufficient large number of polynomial restrictions we can locally approximate any distribution function in the case of classical MaxEnt principle, by using Shannon Entropy. Our derivation of PDF [20]-[22] can be considered as a harmonic analysis under the group of affine transformations. Considering only restrictions with lowest order polynomial means that in the harmonic analysis we restrict ourselves to the lowest dimensional representations.

We will study the stationary point aspect in the MaxEnt problem for the GRE defined by $S_{\mathbf{q}}^{(2)}[\rho]$ for $0 < q_k < 1$ as well as for the domain defined by Eqs.(29) that corresponds to the MaxEnt problem for $S_{\mathbf{p}}^{(1)}[\rho]$, for $p_k > 1$ (See Remark 2).

We will concentrate on the constrained maximization problem in the case $0 < q_k < 1$, $1 \leq k \leq 3$. Recall that we use the notations: $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{q} = (q_1, q_2, q_3)$, $dm(\mathbf{x}) = dm_1(x_1)dm_2(x_2)dm_3(x_3)$, $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$

Denote by $\lambda = (\lambda_0, \dots, \lambda_M)$ the Lagrange multipliers associated to restrictions Eq.(41) and by $\mu(\mathbf{x})$ the multiplier associated to the restriction Eq.(42). From Kuhn-Tucker theorem for maximization [29], we get

$$\frac{\delta}{\delta \rho(\mathbf{x})} \left\{ N_{\mathbf{q}}[\rho] + \int_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x}) \left[\mu(\mathbf{x}) - \sum_{k=0}^M \lambda_k f_k(\mathbf{x}) \right] \right\} = 0 \quad (43)$$

$$\mu(\mathbf{x}) \geq 0; \mu(\mathbf{x}) \rho(\mathbf{x}) = 0 \quad (44)$$

We introduce the following notations

$$g(\lambda, \mathbf{x}) := \frac{1}{q_1 q_2 q_3} \sum_{k=0}^N \lambda_k f_k(\mathbf{x}) \quad (45)$$

$$k_3(\lambda, \mathbf{x}) := [g(\lambda, \mathbf{x})]_+^{1/(q_3-1)} \quad (46)$$

We will use here and in the Appendix 7.2 the following notations for exponents.

$$c_2 = \frac{1 - q_2}{q_2 q_3 - 1} \quad (47)$$

$$d_2 = \frac{q_2(q_3 - 1)}{q_2 q_3 - 1} \quad (48)$$

$$c_1 = \frac{1 - q_1}{q_1 q_2 q_3 - 1} \quad (49)$$

We also denote

$$k_2(\lambda, x_1, x_2) := \int_{\Omega_3} dm_3(x'_3) k_3(\lambda, x_1, x_2, x'_3)^{q_3} \quad (50)$$

$$k_1(\lambda, x_1) := \int_{\Omega_2} dm_2(x'_2) [k_2(\lambda, x_1, x'_2)]^{d_2}; \quad (51)$$

The distribution function that satisfy the stationarity condition Eq.(43) is (for details see Appendix 7.2)

$$\rho(\lambda, \mathbf{x}) = k_3(\lambda, x_1, x_2, x_3) k_2(\lambda, x_1, x_2)^{c_2} [k_1(\lambda, x_1)]^{c_1} \quad (52)$$

5 Examples

5.1 Symmetric distributions

Consider the following examples: $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ with $\Omega_1 = \Omega_2 = \Omega_3 = \mathbb{R}$, $dm_k(x_k) = dx_k$ for $1 \leq k \leq 3$ and for the restrictions from Eq.(41) we have

$$f_0(x_1, x_2, x_3) := 1 \quad (53)$$

$$f_k(x_1, x_2, x_3) := x_k^2; \quad 1 \leq k \leq 3 \quad (54)$$

$$f_k(x_1, x_2, x_3) := x_k^2; \quad 1 \leq k \leq 3$$

By using Eqs.(45-52) we obtain the MaxEnt distribution in the case $0 < q_k < 1$ as follows ($C_{1,2,3}$ are constants)

$$k_2(\lambda, x_1, x_2) = \frac{C_2}{(1 + a_1^2 x_1^2 + a_2^2 x_2^2)^{r_2}} \quad (55)$$

$$r_2 = \frac{q_3}{1 - q_3} - \frac{1}{2} > 0 \quad (56)$$

$$k_1(\lambda, x_1) = \frac{C_1}{(1 + a_1^2 x_1^2)^{r_1}} \quad (57)$$

$$r_1 = n_3 r_2 - \frac{1}{2} > 0 \quad (58)$$

From Eqs.(52, 55, 57) results

$$\rho(\lambda, \mathbf{x}) = a_0 (1 + a_1^2 x_1^2)^{b_1} (1 + a_1^2 x_1^2 + a_2^2 x_2^2)^{b_2} (1 + a_1^2 x_1^2 + a_2^2 x_2^2 + a_3^2 x_3^2)^{-b_3} \quad (59)$$

where a_k are free parameters, and the exponents b_k are given by

$$b_1 = -r_1 n_4 = r_1 \frac{1 - q_1}{1 - q_1 q_2 q_3} \quad (60)$$

$$b_2 = -r_2 n_1 = \left(\frac{q_3}{1 - q_3} - \frac{1}{2} \right) \frac{1 - q_2}{1 - q_2 q_3} \quad (61)$$

$$b_3 = \frac{1}{1 - q_3} \quad (62)$$

The restrictions for $b_{1,2,3}$ resulting from the finiteness of $\langle x_k^2 \rangle$ are

$$b_3 - \frac{3}{2} > 0 \quad (63)$$

$$b_3 - b_2 - 2 > 0 \quad (64)$$

$$b_3 - b_1 - b_2 - \frac{5}{2} > 0 \quad (65)$$

By simple but tedious algebra (we used the Reduce command from MATHEMATICA 5.1) the domain given by Eqs. (56, 58, 60-62), with restriction $0 < q_k < 1$, the resulting domain in the variables q_1, q_2, q_3 is given by

$$1/3 < q_1 < 1 \quad (66)$$

$$\frac{1 + q_1}{4q_1} < q_2 < 1 \quad (67)$$

$$\frac{1 + q_1 + q_1 q_2}{5q_1 q_2} < q_3 < 1 \quad (68)$$

It is easy to verify that the set defined by Eqs(66-68) is not empty: it is an open set and contains at least some open neighborhood of the point defined by $(q_1, q_2, q_3) = (101/152, 267/271, 262/347)$. It can be proven that in the limit $q_k \nearrow 1$ and $a_k \searrow 0$ with suitable scaling we obtain the centered Gaussian distribution, in a similar manner to the following example.

5.2 The RDF for charged particle distribution, derived from Generalized Maximal Entropy principle

5.2.1 The RDF obtained from classical MaxEnt principle [20]-[23]

The reference state studied in [20]-[23] has the form

$$d\hat{\mathcal{F}}^R = \mathcal{N}_0 \left(\frac{w}{\Theta} \right)^{\gamma-1} \exp[-w/\Theta] \exp \left[-c_1 (w/\Theta) (P_\phi - P_{\phi 0})^2 \right] \exp \left[-c_2 (w/\Theta) (\lambda - \lambda_0)^2 \right] |\mathcal{J}| d\hat{\Gamma} \quad (69)$$

where P_ϕ, λ and w are the invariants appearing in the axial symmetric magnetic field variables [20]-[23] Θ is the scale parameter for energy, γ the shape parameter of the gamma distribution that appears in Eq.(69). Supposing that

$$c_1 (w/\Theta) \simeq c_1^{(0)} \equiv \left(\frac{1}{\Delta P_\phi} \right)^2 = \text{const.}$$

$$c_2 (w/\Theta) \simeq c_2^{(0)} + c_2^{(1)} \frac{w}{\Theta} \equiv \frac{1}{\Delta \lambda_0} \left(\frac{\Delta \lambda_0}{\Delta \lambda_1} + \frac{w}{\Theta} \right) \geq 0$$

the density distribution function $\hat{\mathcal{F}}^R$ reads as

$$\hat{\mathcal{F}}^R = \mathcal{N}_0 \left(\frac{w}{\Theta} \right)^{\gamma-1} \exp[-w/\Theta] \exp \left[- \left(\frac{P_\phi - P_{\phi 0}}{\Delta P_\phi} \right)^2 \right] \exp \left[- \left(\frac{\Delta \lambda_0}{\Delta \lambda_1} + \frac{w}{\Theta} \right) \frac{(\lambda - \lambda_0)^2}{(\Delta \lambda_0)^2} \right] |\mathcal{J}| \quad (70)$$

where $\Delta P_\phi, \Delta \lambda_0$ and $\Delta \lambda_1$ are constants and \mathcal{N}_0 ensures normalization to unity.

5.2.2 The RDF obtained from maximal generalized entropy principle

We will study the case when $0 < q_k < 1$, that generate extremal distributions with algebraic decay at infinity. Consider the following restrictions

$$\begin{aligned} f_0(x_1, x_2, x_3) &\equiv 1 \\ f_1(x_1, x_2, x_3) &\equiv |x_1|^{\alpha_1} \\ f_0(x_1, x_2, x_3) &\equiv |x_1|^\delta |x_2|^{\alpha_2} \\ f_0(x_1, x_2, x_3) &\equiv |x_3|^{\alpha_3} \end{aligned}$$

where $x_1 \in \mathbb{R}_+$ and $x_3, x_2 \in \mathbb{R}$. We consider the case of distributions defined in the whole phase space, so we select the Lagrange multipliers strictly positive. According to Eqs.(45-52) as well as Eq.(79) results, up to irrelevant constant

factor

$$\begin{aligned}
k_3(\lambda, \mathbf{x}) &= \left[\lambda_0 + \lambda_1 |x_1|^{\alpha_1} + \lambda_2 |x_1|^\delta |x_2|^{\alpha_2} + \lambda_3 |x_3|^{\alpha_3} \right]_+^{1/(q_3-1)} \\
k_2(\lambda, x_1, x_2) &= \left[\lambda_0 + \lambda_1 |x_1|^{\alpha_1} + \lambda_2 |x_1|^\delta |x_2|^{\alpha_2} \right]^{-m_2} \\
k_1(\lambda, x_1) &= x_1^{-\delta/\alpha_2} [\lambda_0 + \lambda_1 |x_1|^{\alpha_1}]^{-m_1} \\
\rho(\lambda, \mathbf{x}) &= x_1^{-r_1} [\lambda_0 + \lambda_1 |x_1|^{\alpha_1}]^{-r_2} \left[\lambda_0 + \lambda_1 |x_1|^{\alpha_1} + \lambda_2 |x_1|^\delta |x_2|^{\alpha_2} \right]^{-r_3} k_3(\lambda, \mathbf{x})
\end{aligned} \tag{71}$$

where we used the notations

$$m_2 = \frac{q_3}{1-q_3} - \frac{1}{\alpha_3} > 0 \tag{72}$$

$$m_1 = m_2 d_2 - \frac{1}{\alpha_2} > 0 \tag{73}$$

$$r_1 = \frac{\delta}{\alpha_2} c_1; \quad r_2 = m_1 c_1 \tag{74}$$

$$r_3 = m_2 c_2 \tag{75}$$

and Eqs.(47-49). We consider now the choice of the free parameters α_i, δ, q_i in Eq.(52) such that we recover a DDF from the family Eqs.(69, 70). Then according to Eq.(35) we set

$$q_k = 1 - \varepsilon_k; \quad \varepsilon_k \searrow 0 \tag{76}$$

In order to recover Eqs.(69, 70) we set

$$\alpha_1 = 1; \quad \alpha_2 = \alpha_3 = 2$$

without loss of generality we set $\lambda_0 = 1$. Up to linear terms in ε_k we obtain

$$\begin{aligned}
c_2 &= -\frac{\varepsilon_2}{\varepsilon_2 + \varepsilon_3} + \mathcal{O}(\varepsilon_k) \\
d_2 &= \frac{\varepsilon_3}{\varepsilon_2 + \varepsilon_3} + \mathcal{O}(\varepsilon_k) \\
c_1 &= -\frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2 + \varepsilon_3} + \mathcal{O}(\varepsilon_k) \\
m_1 &= \frac{1}{\varepsilon_2 + \varepsilon_3} + \frac{1}{2} \frac{\varepsilon_3}{\varepsilon_2 + \varepsilon_3} + \mathcal{O}(\varepsilon_k) > 0 \\
m_2 &= \frac{q_3}{1-q_3} - \frac{1}{\alpha_3} = \frac{1}{\varepsilon_3} - \frac{3}{2}
\end{aligned}$$

By using the notations

$$\begin{aligned}
s_2 &= \frac{\varepsilon_2 \varepsilon_3}{(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)(\varepsilon_2 + \varepsilon_3)} \\
s_3 &= \frac{\varepsilon_3}{\varepsilon_2 + \varepsilon_3}
\end{aligned}$$

we have the following asymptotic form for the exponents in Eq.(71)

$$\begin{aligned} r_3 &= -\frac{s_3}{\varepsilon_3} + \mathcal{O}(\varepsilon_k) \\ r_2 &= -\frac{s_2}{\varepsilon_3} + \mathcal{O}(\varepsilon_k) \\ r_1 &= -\frac{\delta\varepsilon_1}{\alpha_2(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)} + \mathcal{O}(\varepsilon_k) \end{aligned}$$

Now we consider the limit $\varepsilon_k \rightarrow 0$, with $\lambda_0 = 1$ and s_2, s_3, r_1 fixed and $\lambda_k \rightarrow 0$ such that

$$\lambda_k = \varepsilon_3 \nu_k; 1 \leq k \leq 3$$

and $\nu_k; 1 \leq k \leq 3$ are fixed positive constants. We obtain the following result in the limit $\varepsilon_3 \rightarrow 0$

$$\rho(\lambda, \mathbf{x}) = x_1^{-r_1} \exp[-\mu_1 x_1 - \mu_2 x_1^\delta x_2^2 - \mu_3 x_3^2] \quad (77)$$

where the following notation was used

$$\begin{aligned} \mu_1 &= \nu_1 \frac{\varepsilon_2}{(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)} > 0 \\ \mu_2 &= \nu_2 \frac{\varepsilon_2}{(\varepsilon_2 + \varepsilon_3)} > 0 \\ \mu_3 &= \nu_3 > 0 \end{aligned}$$

In order to approach the class of RDF from Eqs.(69-70), we put in Eq.(77)

$$\begin{aligned} x_1 &= w \\ x_2 &= \lambda - \lambda_0 \\ x_3 &= P_\Phi - P_{\Phi 0} \\ r_1 &= 1 - \gamma < 1 \\ \frac{1}{\Theta} &= \mu_1 \\ \frac{1}{(\Delta P_\Phi)^2} &= \mu_3 \end{aligned}$$

and we obtain the limiting RDF of the form [24]

$$\hat{\mathcal{F}}^R = \mathcal{N}_0 \left(\frac{w}{\Theta} \right)^{\gamma-1} \exp[-w/\Theta] \exp \left[- \left(\frac{P_\phi - P_{\phi 0}}{\Delta P_\phi} \right)^2 \right] \exp \left[-w^\delta \frac{(\lambda - \lambda_0)^2}{(\Delta \lambda_0)^2} \right] \quad (78)$$

which excepting to a constant factor in $\frac{(\lambda - \lambda_0)^2}{(\Delta \lambda_0)^2}$ term reproduces well the qualitative behavior of the RDF from Eqs.(69-70).

However, it is correct to mention that the RDFs (69, 70) and Eq. (78) should be understood as pure *pedagogical* distribution functions and not as realistic RDFs. Indeed, it is easily to check that there exist several ranges of parameters where

these RDFs violate the energy balance equations. The interest of these RDFs resides solely on the fact that they can be used as RDF-fitting functions and that, as demonstrated in this work and in [20]-[23], these fittings are derivable from the MaxEnt principle.

6 Conclusions

The nonlinear equations associated to the maximal Generalized Rényi Entropy problem was solved in the case of GRE with 3 variables. By simple linear scale invariant restrictions a family of reference particle distribution function for charged particles in tokamak was obtained.

7 Appendix

7.1 Integrals

The following formula will be used

$$\int_0^\infty \frac{dx}{(a + bx^\alpha)^r} = b^{-1/\alpha} a^{(1/\alpha - r)} \frac{1}{\alpha} B(1/\alpha, r - 1/\alpha) \quad (79)$$

where $a > 0$, $\alpha > 0$, $b > 0$, $r > 1/\alpha$.

Proof. Eq.(79) is reduced to the standard integral representation of the Euler Beta function by the substitution

$$x = \left(\frac{a}{b} \frac{u}{1-u} \right)^{1/\alpha}$$

■

7.2 Derivation of the extremal PDFs

From Eqs.(43, 45) we obtain the following nonlinear integral equation for PDF $\rho(x_1, x_2, x_3)$

$$B(x_1)^{q_1-1} A(x_1, x_2)^{q_2-1} \rho(x_1, x_2, x_3)^{q_3-1} = g(\lambda, \mathbf{x}) \quad (80)$$

where we denoted

$$A(x_1, x_2) = \int_{\Omega_3} dm_3(x'_3) \rho(x_1, x_2, x'_3)^{q_3} \quad (81)$$

$$B(x_1) = \int_{\Omega_2} dm_2(x'_2) [A(x_1, x'_2)]^{q_2} \quad (82)$$

In the following we will use the notations from Eqs.(45-51). The solution of system of Eqs (80-82) is of the form

$$\rho(x_1, x_2, x_3) = \rho_2(x_1, x_2)g(\lambda, \mathbf{x})^{1/(q_3-1)} = \rho_2(x_1, x_2)k_3(\lambda, \mathbf{x}) \quad (83)$$

where $\rho_2(x_1, x_2)$ remains to be determined. By inserting Eq.(83) in Eqs.(80, 81) we obtain successively (see notations from Eqs.(50, 51))

$$B(x_1)^{q_1-1}A(x_1, x_2)^{q_2-1}\rho_2(x_1, x_2)^{q_3-1} = 1 \quad (84)$$

$$A(x_1, x_2) = \rho_2(x_1, x_2)^{q_3}k_2(\lambda, x_1, x_2) \quad (85)$$

We will use the notations for exponents

$$c_1 = \frac{1 - q_1}{q_1 q_2 q_3 - 1} \quad (86)$$

$$d_2 = \frac{q_2(q_3 - 1)}{q_2 q_3 - 1} \quad (87)$$

$$c_2 = \frac{1 - q_2}{q_2 q_3 - 1} \quad (88)$$

$$n_2 = \frac{q_3 - 1}{q_2 q_3 - 1} \quad (89)$$

From Eqs.(84, 85) results

$$\rho_2(x_1, x_2) = \rho_1(x_1)k_2(\lambda, x_1, x_2)^{c_2} \quad (90)$$

$$A(x_1, x_2) = \rho_1(x_1)^{q_3}k_2(\lambda, x_1, x_2)^{n_2} \quad (91)$$

where $\rho_1(x_1)$ remains to be found. Now we insert Eqs.(91, 90) in Eqs.(82, 84), we obtain

$$B(x_1) = \rho_1(x_1)^{q_2 q_3}k_1(\lambda, x_1) \quad (92)$$

$$B(x_1)^{q_1-1}\rho_1(x_1)^{q_2 q_3-1} = 1 \quad (93)$$

and from here we get

$$\rho_1(x_1) = [k_1(\lambda, x_1)]^{c_1} \quad (94)$$

Collecting the previous results we obtain

$$\rho(\lambda, \mathbf{x}) = k_3(\lambda, x_1, x_2, x_3)k_2(\lambda, x_1, x_2)^{c_2} [k_1(\lambda, x_1)]^{c_1} \quad (95)$$

References

- [1] C. E. Shannon, *A Mathematical Theory of Communication, Bell System Technical Journal*, **27**, 379-423 & 623-656 (1948).
- [2] C. Tsallis, *Journal of Statistical Physics*, **52**, 479-487 (1988).

- [3] C. Tsallis, R. S. Mendes and A. R. Plastino, *Physica A*, **261**, 534-554 (1998).
- [4] C. Tsallis, *Introduction to nonextensive statistical mechanics_ approaching a complex world*, (Springer, 2009).
- [5] M. Gell-Mann, C. Tsallis, *Nonextensive entropy- Interdisciplinary applications*, (Oxford University Press, USA, 2004).
- [6] A. Rényi (1960), *On measures of information and entropy*, *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, June 20-July 30, 1960, Volume I, University of California Press, Berkeley and Los Angeles, 547-561 1960*. pp. 547–561.
- [7] A. Rényi, *Wahrscheinlichkeitstheorie* (De Gruyter, Berlin, 1974).
- [8] A. Rényi, *On statistical laws of accumulation of information*, in *Rényi Alfréd Válogatott Munkái (A. Rényi Selected works)*, vol. **3**, 1962-1970, Akadémiai Kiadó, Budapest 1977, page 33.
- [9] A. Rényi, *Rev. Int. Inst. Stat.*, **33**, 1-14 (1965).
- [10] Yu. L. Klimontovich, *Chaos, Solitons, Fractals*, **5**, 1985-2002 (1995).
- [11] I. Csiszár, *I.E.E.E. Transactions on Information Theory*, **41** -1, 26, (1995).
- [12] T. van Erven, P. Harremoës, "Rényi Divergence and Kullback-Leibler Divergence", e-print arXiv:1206.2459v2 (2014).
- [13] M. Reed and B. Simon (1981), *Functional Analysis (Methods of Modern Mathematical Physics)*. Vol. **1**, Academic Press.
- [14] W. Rudin (1987), *Real and Complex Analysis*; McGraw Hill Inc. 3rd Ed. page 74.
- [15] H. Luschgy and G. Pagès, *Moment estimates for Lèvy processes. Electronic Communication in Probability*, **13**, 422-434 (2008).
- [16] G. Sonnino, G. Steinbrecher, *Phys. Rev. E* **89**, 062106 (2014).
- [17] G. Steinbrecher, G. Sonnino, *Numerical Stability of Generalized Entropies*, arXiv:1603.06240v3 [physics.data-an] (2016).
- [18] O. V. Besov, V. P. Il'in and S.M. Nikol'skii (1975), *Integral representations of functions and embedding theorems*, Ed. Nauka, pp. 9-40, Moscow (in Russian).
- [19] G. Steinbrecher, A. Sonnino, A. and G. Sonnino, *Journal of Modern Physics*, **7**, 251-266 (2016).
- [20] G. Sonnino, G. Steinbrecher, A. Cardinali, A. Sonnino, M. Tlidi, *Phys. Rev. E*, **87**, , 014104-1, 014104-5 (2013).

- [21] G. Sonnino, A. Cardinali, G. Steinbrecher, P. Peeters, A. Sonnino, P. Nardone, *Physics Letters A*, **377**, 3061-3077 (2013).
- [22] G. Sonnino, A. Cardinali, P. Peeters, G. Steinbrecher, A. Sonnino, P. Nardone, *The European Physical Journal D*, **68**, 1 - 11 (2014) .
- [23] G. Sonnino, P. Peeters, A. Sonnino, P. Nardone, G. Steinbrecher, *Journal of Plasma Physics*, **81**, 01-905810116. (2015).
- [24] C. Di Troia, *Plasma Physics and Controlled Fusion* (PFCF), **54**, 105017 (2012).
- [25] G. Steinbrecher, G. Sonnino, *Generalized Rényi Entropy and Structure Detection of Complex Dynamical Systems*, arXiv:1512.06108v1 [physics.data-an] (2015).
- [26] E. M. Stein, R. Shakarchi, *Functional Analysis, Introduction to further topics, Princeton Lectures in Analysis IV, Princeton Univ. Press, Princeton (2011),page 36 exerc. 9.c. page 39 exerc. 20.*
- [27] E. M. Stein, *Singular Integrals and Differentiable Properties of Functions*. Princeton Univ. Press. (1971).
- [28] S. Boyd and L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press, Cambridge.
- [29] S. Stephen and L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press. Cambridge, page 244.
- [30] G. Sonnino, *Eur. Phys. J. D*, **62**, 81 (2011).